DOCUMENT RESUME

ED 028 800 LI 001 457

By-Buckland, Lawrence F.; And Others

Demonstration of Cataloging Support Services and Marc II Conversion. Final Report.

Inforonics, Inc., Cambridge, Mass.

Spons Agency-Council on Library Resources, Inc., Washington, D.C.; New England Board of Higher Education, Wellesley, Mass.

Report No-CLR-425

Pub Date 2 Jan 69

Note-40p.

EDRS Price MF-$0.25 HC-$2.10

Descriptors-Automation, *Cataloging, Centralization, Computer Programs, Demonstration Projects, *Electronic Data Processing, *Information Processing, *Library Networks, Library Services, *Library Technical Processes, Pilot Projects

Identifiers-Machine-Readable Cataloging, MARC, NELINET, *New England Library Information Network

Beginning in December, 1967, the New England Library Information Network (NELINET) was demonstrated in actual operation using Machine-Readable Cataloging (MARC I) bibliographic data. Section 1 of this report is an introduction and summary of the project. Section 2 described the library processing function demonstrated which included catalog card and label services. The early months of the project were devoted to solving problems with the system. During May, June, and July, 1968, attention was concentrated on achieving a more efficient pilot operation. As part of this effort, statistics were compiled in June and July. From these statistics an estimate is made of cost per title of performing a similar operation on a full scale random access system. Appendix A contains this cost projection. The demonstration of cataloging services was suspended on July 31, 1968, and the project was redirected to setting up a MARC II based system. Section 3 describes this effort. The basic difficulty was deciding whether immediate hook-up with interim programs or delayed hook-up with permanent programs was better. The decision was made in favor of delayed hook-up and programs suiting this system are described. (Author/CC)

ED028800

Inforonics*

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

LI 001457

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

NELINET - NEW ENGLAND LIBRARY INFORMATION NETWORK

DEMONSTRATION OF CATALOGING SUPPORT SERVICES AND MARC II CONVERSION

PREPARED BY

LAWRENCE F. BUCKLAND, ANN T. CURRAN, WILLIAM R. NUGENT

SUBMITTED TO

THE NEW ENGLAND BOARD OF HIGHER EDUCATION

FINAL REPORT

CONTRACT NO. CLR-425

LI 001457

JANUARY 2, 1969

ERIC/CLIS
MAY 29 '69
UNIVERSITY OF MINNESOTA

17

*new concepts in information organization, processing, and presentation

INFORONICS, INC.
806 MASSACHUSETTS AVENUE, CAMBRIDGE, MASSACHUSETTS 02139, TEL. (617) 547-1750
146 MAIN STREET, MAYNARD, MASSACHUSETTS 01754, TEL. (617) 897-8815
927 15TH STREET, N. W., WASHINGTON, D. C. 20005, TEL. (202) 638-6862

TABLE OF CONTENTS

		PAGE
1.	INTRODUCTION	1
2.	DEMONSTRATION OF SERVICES	2
	2.1 PROCEDURE	2
	2.2 PRODUCTION STATISTICS	12
	2.2.1 PRODUCTION SUMMARIES	12
	2.2.2 PRODUCTION SUMMARIES BY LIBRARY	14
	2.2.3 TURN AROUND TIMES	16
	2.2.4 PERCENT OF CARD SETS GENERATED PER RUN	17
	2.2.5 MACHINE RUNNING TIMES	17
	2.2.6 MACHINE RUNNING COSTS	17
	2.3 PILOT DEMONSTRATION SUSPENDED JULY 31	17
3.	CONVERSION TO THE MARC II BASED SYSTEM	21
	3.1 INTRODUCTION	21
	3.2 SYSTEMS DEVELOPMENT AND PROGRAMMING	24
	3.2.1 MASTER FILE GENERATOR	25
	3.2.2 LC MARC II TO NELINET MARC II CONVERTER	26
	3.2.3 SEARCH/MERGE	27
	3.2.4 CARD AND LABEL PRODUCTION	27
	3.2.5 CARD FORMATTER	29
4.	CONCLUSION	31

LIST OF ILLUSTRATIONS

		PAGE
FIGURE 1	DEMONSTRATION FLOW CHART	3
FIGURE 2	REQUEST WORKSHEET	4
FIGURE 3	TELETYPE REQUEST	5
FIGURE 4	SELIN LABELS	7
FIGURE 5	CATALOG CARDS	9
FIGURE 6	CATALOG CARDS	10
FIGURE 7	PROBLEM REPORT	11
TABLE I	PRODUCTION SUMMARIES	13
TABLE II	PRODUCTION SUMMARIES BY LIBRARY	15
TABLE III	TURN AROUND TIMES	16
TABLE IV	PERCENT OF CARD SETS GENERATED PER RUN	18
TABLE V	MACHINE RUNNING TIMES	19
TABLE VI	MACHINE RUNNING COSTS	20

1. INTRODUCTION AND SUMMARY

The library system designed in the project of 1966, and developed and tested at the University of New Hampshire in 1967, became a reality in spring of 1968 when NELINET became an operating library network. Between spring and fall of 1968 the four libraries of the University of Connecticut, the University of Vermont, the University of Rhode Island and the University of Massachusetts were linked into the system. In May, June, and July the network was demonstrated in actual operation using MARC I bibliographic data.

The library processing function demonstrated operated as follows: (1) the libraries transmitted up to 40 requests for catalog cards to the processing center by teletype twice a week, (2) the requests were run through a series of computer programs which searched MARC tapes and, for those titles found, produced a magnetic tape containing catalog card images, and a paper tape containing the Selin label images, (3) the next morning the Library of Congress card numbers for the titles that were not found were transmitted to the libraries, (4) the paper tape was printed out on a tape typewriter to produce the Selin labels, (5) the magnetic tape was run on the 1403 line printer at Widener Library to produce catalog card sets, (6) the output products, cards and labels, were reviewed by Inforonics' staff and mailed to the libraries the following day.

As anticipated, many problems were encountered under this system. In the early months of the demonstration Inforonics and the five libraries devoted their efforts to solving these problems. During May, June, and July attention was concentrated on achieving a more efficient pilot operation. As part of this effort, statistics were compiled in June and July. From these statistics an estimate is made of cost per title of performing a similar operation on a full scale random access system. Appendix A contains this cost projection.

The demonstration of cataloging services was suspended on July 31 and the project was redirected to setting up a MARC II based system. Section 3 describes this effort; the basic difficulty was deciding whether immediate hook-up with interim programs or delayed hook-up with permanent programs was better. The decision was made in favor of delayed hook-up; and programs suiting this system are described.

2. DEMONSTRATION OF SERVICES

Demonstration of catalog card and label services began in December 1967 with the University of New Hampshire participating during the shake down period of December through March. With the transmittal of requests by the University of Rhode Island on April 1, NELINET began its transition from a one library test situation to an operating network. With the addition of the University of Massachusetts on April 24, the University of Connecticut on April 30, and the University of Vermont on May 7, the network was operating in all the participating libraries. (Maine withdrew during the month of April, 1968.)

2.1 PROCEDURE

Prior to accepting any of the above libraries into the network, each library's staff was instructed in the use of the service. These instructions have been supplemented by teletype or telephone communications when incorrect requests were received. The system operating procedure is as follows: (the letter of each step corresponds to the letter in the flow chart, Figure 1.)

(a) The cataloger (or catalog assistant) fills out a request worksheet for titles expected to be on the MARC I tapes (current titles in English). This worksheet (see Figure 2) contains the library's symbol, the Library of Congress card number, the local copy, volume and branch information, and the call number if the library does not desire the one established at the Library of Congress. The libraries can also request extra copies of the main entry or suppress labels or catalog cards if they wish.

(b) The teletype operator types the information recorded on the worksheet along with the library code. (See Figure 3.) In order to distinguish upper and lower case characters, "$" precedes every upper case character. These messages are typed and corrected off-line and then transmitted to Inforonics twice a week on Monday and Wednesday mornings. Each library can submit up to 40 requests. The teletype at Inforonics produces a punched paper tape and a listing of each library's requests.

(c) The "Request Validation Program" validates the paper tape requests for particular machine detectable errors, converts the data into master file format and the character codes into master file

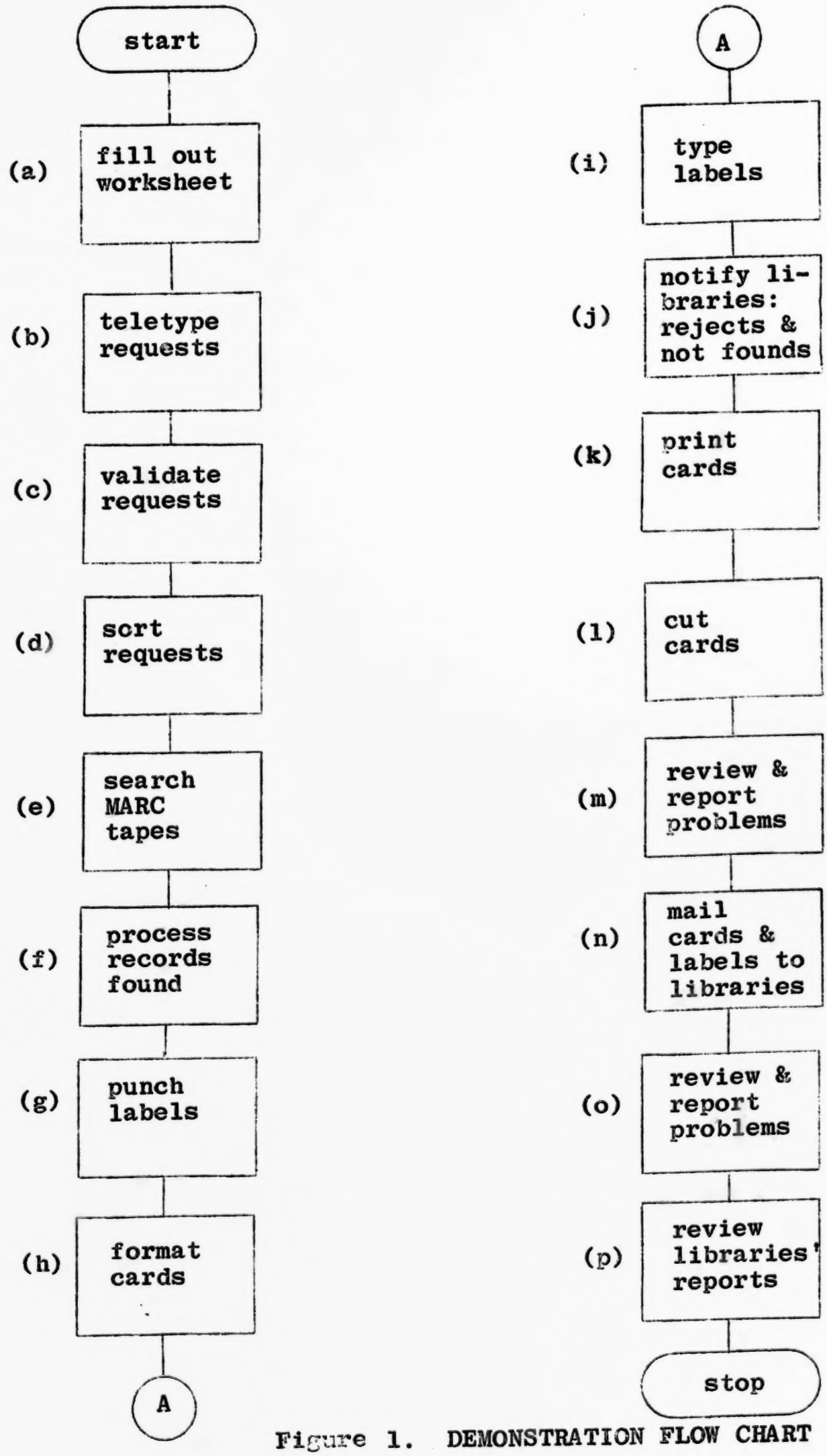


Figure 1. DEMONSTRATION FLOW CHART

NELINET Request Form

MA

LC ☐ 66-11631
L.C. Card Number

AT ☐

Location	Copy-Shelf Statement	Volume
AT ☐ Main;	1c.1-2;	1v.1-3
AT ☐ Chem;	c.1	1v.1-3
AT ☐		

CN ☐ Local Call Number

XC ☐ Extra Main Entry Cards

N C L P
No Cards Labels Pockets

\\

Branches	Shelf Locations and Branches without Catalogs
AG EN	BUS
CHEM	NUR
CRAN	PSYCH
EDUC	RES C
ENGIN	TECH P
ENT	+
FOOD	FOLIO
FOR	FFOLIO
GOV R	OFF
HOME	REF
LABOR	PER
LAND	
MATH	
MORR	
MUSIC	
PHYS	
PLANT	
SHADE	
VET	
WALT	
SPEC	

Figure 2: REQUEST WORKSHEET

```
MA

LC 66-11631

AT MAIN; ↑ C.1-2; ↑ V.1-3

AT SCSHSEEM; C.1: ↑ V.1-3


\\
```

Figure 3: TELETYPE REQUEST

character codes, and outputs all valid requests onto magnetic tape. It also outputs messages regarding those requests that had been rejected because of parity errors or invalid data.

(d) The "Request Sort Program" sorts the requests by Library of Congress card number into one numerical sequence.

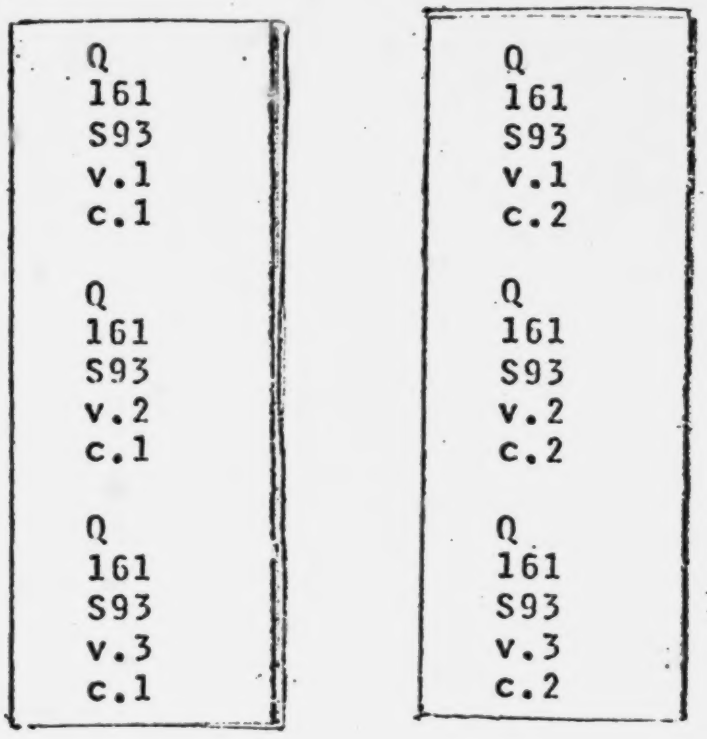
(e) The "Search Program" matches the requests against the MARC tapes which are also in card number order. Records that are found are written onto another tape, and records that are not found are listed. This "search program" also performs some of the card processing functions that would have been included in the "Card Production Program" had the PDP-1 been large enough to accommodate all functions necessary in one program.

(f) The "Card Production Program" processes the output tape from the search program. This program includes a profile for each library. Information about the number of cards needed for branch books and the dimensions and symbols each library uses to distinguish and locate oversize books, is contained in this profile. Using the library's profile, the data in the request and in the MARC record is repeatedly duplicated onto another tape so that there is a separate record for every card and label required.

(g) The "Punch Labels Program" punches a paper tape in DURA character codes from which Selin labels are produced.

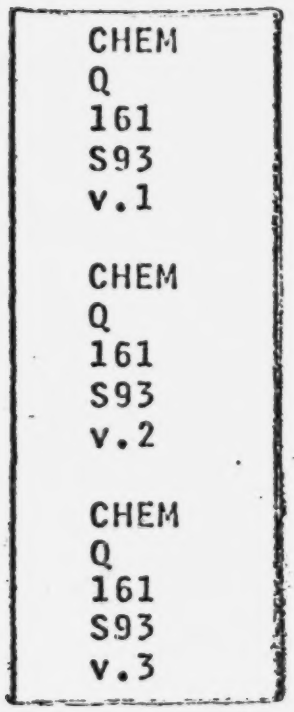
(h) The "Card Formatter Program" formats the records produced by the "Card Production Program," converting the master file character codes into 1403 character codes and generating continuation card headers when necessary.

(i) The paper tape from the "Punch Labels Program" is run on the DURA tape typewriter and Selin labels are typed. (See Figure 4.)



SELIN LABELS FOR VOLUMES IN

MAIN LIBRARY



SELIN LABELS FOR VOLUMES IN

CHEMISTRY LIBRARY

Figure 4: SELIN LABELS

(j) The librarians are notified the next morning of the requests that were rejected and the records that were not found on the MARC tape. This enables the libraries to return these books to be processed by their manual system with a minimum of delay.

(k) The magnetic tape from step h is taken to Widener Library at Harvard University where catalog cards are printed on continuous card stock by a 1403 line printer with upper and lower case characters. (See Figures 5 and 6.)

(l) The catalog cards are then mechanically cut by a NIKOR card cutter.

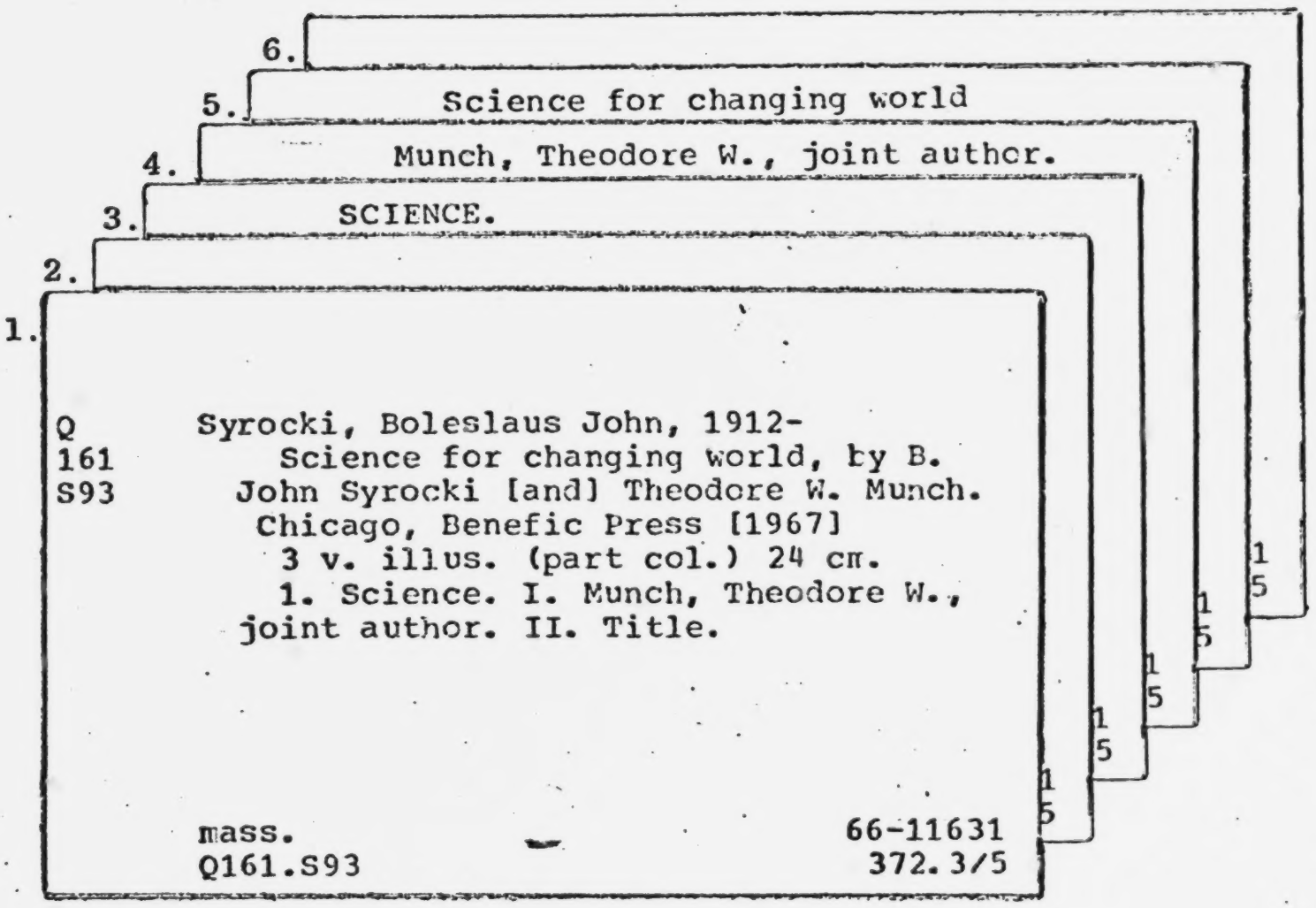
(m) A librarian at Inforonics reviews all cards and labels and fills out a problem report if any program bugs are detected. (During the demonstration all cards are sent out on the assumption that the libraries would be interested in seeing even the unfileable ones.) (See Figure 7.)

(n) The cards and labels are mailed to the libraries.

(o) The libraries review the cards and labels and fill out problem reports which they return to Inforonics.

(p) Inforonics' staff reviews the problem reports. Often these problems involve preferences in style rather than program bugs; e.g., the library would prefer to have all the tracings printed on one card rather than have them begin on one card and continue on the next. Copies of the problem reports, those submitted by Inforonics' staff as well as by the libraries, are sent to the Council on Library Resources and to the members of the Advisory Group who request them.

Card and label shipments were few and none-too-prompt in the beginning stages. The planned turn-around-time (the time from request to shipment) of two days was not consistently met until the demonstration had been running for some months.



MAIN CATALOG

(VOLUMES IN MAIN LIBRARY)

Figure 5: CATALOG CARDS

11.
10. Science for changing world
9. Munch, Theodore W., joint author.
8. SCIENCE.
7.

CHEM
Q
161
S93

Syrocki, Boleslaus John, 1912-
Science for changing world, by B.
John Syrocki [and] Theodore W. Munch.
Chicago, Benefic Press [1967]
3 v. illus. (part col.) 24 cm.
1. Science. I. Munch, Theodore W.,
joint author. II. Title.

mass.
Q161.S93

66-11631
372.3/5

CHEMISTRY CATALOG

(VOLUMES IN CHEMISTRY LIBRARY)

18.
17.
16. Science for changing world
15. Munch, Theodore W., joint author.
14. SCIENCE.
13.
12.

CHEM
Q
161
S93

Syrocki, Boleslaus John, 1912-
Science for changing world, by B.
John Syrocki [and] Theodore W. Munch.
Chicago, Benefic Press [1967]
3 v. illus. (part col.) 24 cm.
1. Science. I. Munch, Theodore W.,
joint author. II. Title.

mass.
Q161.S93

66-11631
372.3/5

MAIN CATALOG

(VOLUMES IN CHEMISTRY LIBRARY)

Figure 6: CATALOG CARDS

68-18

OMICS NELINET -11-

Library Services Report

Problem or Suggestion

1. From: J. Agenbroad 2. Date: Jan. 15, 1968
mo. day year

3. To: R. Simmons

4. Description of problem: (attach sample if possible)
Present programs conider a subject heading subdivision as an extension of the last word in the main heading because the double dashes are characters. This makes for very wide right margins and occasionally, an unnecessary carry over onto card 2.

5. Suggested Improvement: Have the double dashes treated as a valid breaking point for a line, before or after the "--". It would be desirable and probably easier to regard these dashes as a valid break whenever they occur (overprint heading, contents note, etc.)

6. Diagnosis:(describe any attempts to fix)

7. Suggested remedy:

LS 2 January, 1968

Figure 7: PROBLEM REPORT

2.2 PRODUCTION STATISTICS

During the last two months of the demonstration when it was running fairly smoothly, quantitative statistics were compiled pertinent to production volume, machine running times, and the timeliness of the service. In summary, 2537 requests were received for which 1149 MARC records were found and 1020 acceptable card sets were made. These statistics were a necessary step in analyzing production, and were needed by all those involved in the project in order to assess the performance of the test system, and to project this assessment to a future system with full production capacity.

2.2.1 Production Summaries

The total volume of all test production is summarized in Table I. In the beginning months Inforonics' staff manually proofread and corrected requests. Although the amount of data to be keyed was small, some time was needed to become familiar with the teletypewriter and the request format. Some mistakes such as an LC card number with the year missing were uncorrectable and were excluded from the computer run. Inforonics stopped correcting mistakes in July. Validation routines had been programmed into the request processing program to catch likely errors that were machine detectable. When errors were present in a request, Xerox copies of the teletype request hard copy were made, errors were noted, and the Xerox copies were returned to the libraries.

Included in the number of rejects is also the number of records that were rejected because of parity errors in the teletype paper tape. As a rule, more records were rejected because of parity errors than library errors. The increased number of rejects in July was the result of hardware parity errors.

In June, 51.9% of the requests were not found on the MARC I tapes whereas only 39% were not found in July. Since the Library of Congress stopped inputting at the end of June, it was expected that the number of "not founds" would gradually increase. However, this was not the case, and some factors which help to explain why follow. First, the University of New Hampshire was unable to have their teletypewriter fixed because of the telephone strike. Thus, the University of New Hampshire, which always had a high per cent of "not founds," was unable to submit requests during July. Second, some of the libraries had, through experience, gained a better idea of what to expect on the tapes and used some selection principles before requesting. In June, 40.6% of the University of Massachusetts' requests were not found whereas only 22.1% were not found in July.

The number of card sets missing warrants attention and some explanation. In a full scale production operation, such performance could not be allowed. Card sets (for titles that are on the MARC tape) may be missing for three reasons: (1) a program bug; (2) a hardware malfunction; or (3) an oversized record.

TABLE I

PRODUCTION SUMMARIES

	June		July	
	Number	Per Cent*	Number	Per Cent*
Total Requests	1056		1481	
Requests Uncorrectable	4	.4	7	0.4
Requests Processed	1052	99.6	1474	99.6
Requests Computer Rejected	36	3.4	90	6.1
Requests Searched	1016	96.2	1384	93.5
Requests Not Found	548	51.9	577	39.0
Card Sets Missing	27	2.6	99	6.7
Card Sets Produced	441	41.7	708	47.8
Card Sets Acceptable	398	37.7	622	42.0
Card Sets Unacceptable	43	4.0	86	5.8

* All percentages are of total

The programs required to produce catalog cards are long and complicated, incorporating several thousand individual instructions. The presence of errors in such programs is expected and the need for debugging accepted. What may be underestimated is the amount of "testing" or "running" that is required to completely debug such programs. Cataloging data varies considerably from one record to another. Before all possible conditions and combinations of conditions have been met, many thousands of records will have to be run, not 500, or 1000, or 2000.

During the month of July, requests were resubmitted for the missing card sets of previous runs. Many of these, 31.9%, were generated in the second run. Although it might appear that hardware malfunction was the reason the card set was missing from the original run, this might not always be the case. Possibly the conditions present in a record preceding the missing one in the original run were the reason that the missing one was not generated. To attribute with certainty the cause of the missing card set to hardware malfunction would require rerunning the original batch of requests. In the MARC II pilot demonstration, extensive bug detecting and exterminating efforts will be required since this will be the eventual operating system. Consideration is also being given in the MARC II system to developing a more production oriented system, one which will monitor itself; such a system will be easier to obtain within a larger machine configuration.

In a larger machine system, it will be possible to process large records. The amount of core available in the PDP-1 computer is not large enough to accommodate both the long programs required to generate catalog cards and the work area required to process large catalog records.

The card sets reported as unacceptable include those with some program bug, those with keying errors in the MARC data, and those ruined by the card cutter.

2.2.2 Production Summaries by Library

The production volume for each library is summarized in Table II. Although one would have expected that the percentage of records found on the MARC tape would not have varied significantly from library to library, the opposite was found to be true. The reason for such variation is that the libraries inserted their requesting procedures at different times in the processing cycle. The University of Connecticut, the library with the highest percentage of "not founds," requested cards immediately upon receipt of the book. Most of the receipts in this period were for standing orders. These orders were sent immediately upon publication of the item and were received at the University of Connecticut before the Library of Congress had them processed and on the MARC tapes. The University of Vermont, on the other hand,

TABLE II

PRODUCTION SUMMARIES BY LIBRARY

	Conn.		Mass.		N.H.		R.I.		Vt.	
	Je	Jy	Je	Jy	Je	Jy	Je	Jy	Je	Jy
Total Requests	194	248	281	426	142	0	238	405	201	402
Requests Uncorrectable	0	2	0	1	0	0	0	1	4	3
Requests Processed	194	246	281	425	142	0	238	404	197	399
Requests Computer Rejected	9	12	5	32	13	0	4	20	5	26
Requests Searched	185	234	276	393	129	0	234	384	192	373
Requests Not Found	162	179	114	94	85	0	136	211	52	93
Card Sets Missing	1	5	1	52	8	0	13	18	3	24
Card Sets Produced	22	50	161	247	36	0	85	155	137	256
Card Sets Acceptable	21	46	150	219	29	0	72	138	125	219
Card Sets Unacceptable	1	4	11	28	7	0	13	17	12	37

PRODUCTION SUMMARIES BY LIBRARY--PER CENT*

	Conn.		Mass.		N.H.		R.I.		Vt.	
	Je	Jy	Je	Jy	Je	Jy	Je	Jy	Je	Jy
Total Requests										
Requests Uncorrectable	0.0	0.6	0.0	0.3	0.0	0	0.0	0.3	2.0	0.7
Requests Processed	100.0	99.4	100.0	99.7	100.0	0	100.0	99.7	98.0	99.3
Requests Computer Rejected	4.6	4.8	1.8	7.5	9.2	0	1.7	4.9	2.5	6.5
Requests Searched	95.4	94.6	98.2	92.2	90.8	0	98.3	94.8	95.5	92.8
Requests Not Found	83.5	72.2	40.6	22.1	59.9	0	57.2	52.0	25.8	23.1
Card Sets Missing	0.5	2.0	0.4	12.2	5.6	0	5.5	4.5	1.5	6.0
Card Sets Produced	11.4	20.4	57.2	57.9	25.3	0	35.6	38.3	68.2	63.7
Card Sets Acceptable	10.9	18.8	53.5	51.3	20.4	0	30.1	33.8	62.2	54.5
Card Sets Unacceptable	0.5	1.6	3.9	6.6	4.9	0	5.5	4.5	6.0	9.2

* All percentages are of total

did not request catalog cards until they were ready to process the book, generally some time after its publication. They experienced a low percentage of "not founds."

2.2.3 Turn Around Times

TABLE III

	No. of shipments reported			No. of (working) days from request by library to receipt of shipment			
LIBRARY	JUNE	JULY	TOTAL	3 DAYS	4 DAYS	5 DAYS	6 DAYS
Conn.	5	6	11	6	5		
Mass.	6	9	15	8	2	5	
N. H.	4	0	4	2			2
R. I.	3	10	13	8	2	3	
Vt.	2	9	11	9		2	

The schedule aimed for by Inforonics was a two-day turn around time. Requests received on one morning would be run on the PDP-1 the same day; the paper tape output for Selin labels would be run on the DURA machine in Maynard and the magnetic tape containing the catalog card images would be sent to Inforonics' Cambridge office the same evening. The magnetic tape would be run on the 1403 line printer at Harvard the next day. Since this was usually scheduled late in the day, the catalog cards generated would be reviewed the following day and mailed to the libraries on that day. If this schedule was met, catalog cards for requests made by the libraries on Monday morning would be mailed on Wednesday and those requests made on Wednesday would be mailed on Friday.

A total of seventeen runs was made and reported on in June and July. Of these, eleven were mailed two days after request; one run was mailed one day after request; four were mailed three days after request; and one was mailed five days after request. Of the four runs mailed three days after request, two were delayed because the line printer was not operating at Widener; one was delayed because of Inforonics' staffing inadequacies; and one was delayed because of the July 4th holiday. The malfunctioning of the card cutter caused the five day turn around on July 31, the last day of the demonstrations.

Table III should be interpreted as follows: of the 11 shipments reported by the University of Vermont, 9 were received 3 working days after they were requested and 2 were received 5 working days after they were requested.

2.2.4 Percent of Card Sets Generated Per Run

As can be seen in Table IV, the percentage of titles found on the MARC tapes varied from day to day. The greatest change was from 14% on June 5 to 44% on June 10. The increase is explained by the fact that the MARC data file was updated by the receipt of additional records.

2.2.5 Machine Running Times

The actual computer time used for the separate processes are summarized in Table V. The first six programs listed in Table V were run on the PDP-1 computer. The last one was run on the 1403 line printer at Harvard.

In July one less tape drive was available causing the increase in searching time.

2.2.6 Machine Running Costs

The computer running times were converted to costs and are summarized in Table VI. The printing of cards on the 1403 line printer was the most costly step in the process. This suggests the desirability of printing two cards across at a time.

Searching was the next most expensive step. Searching costs for this experimental demonstration are not too meaningful since they will vary directly with the size of the file being searched and inversely with the number of records found.

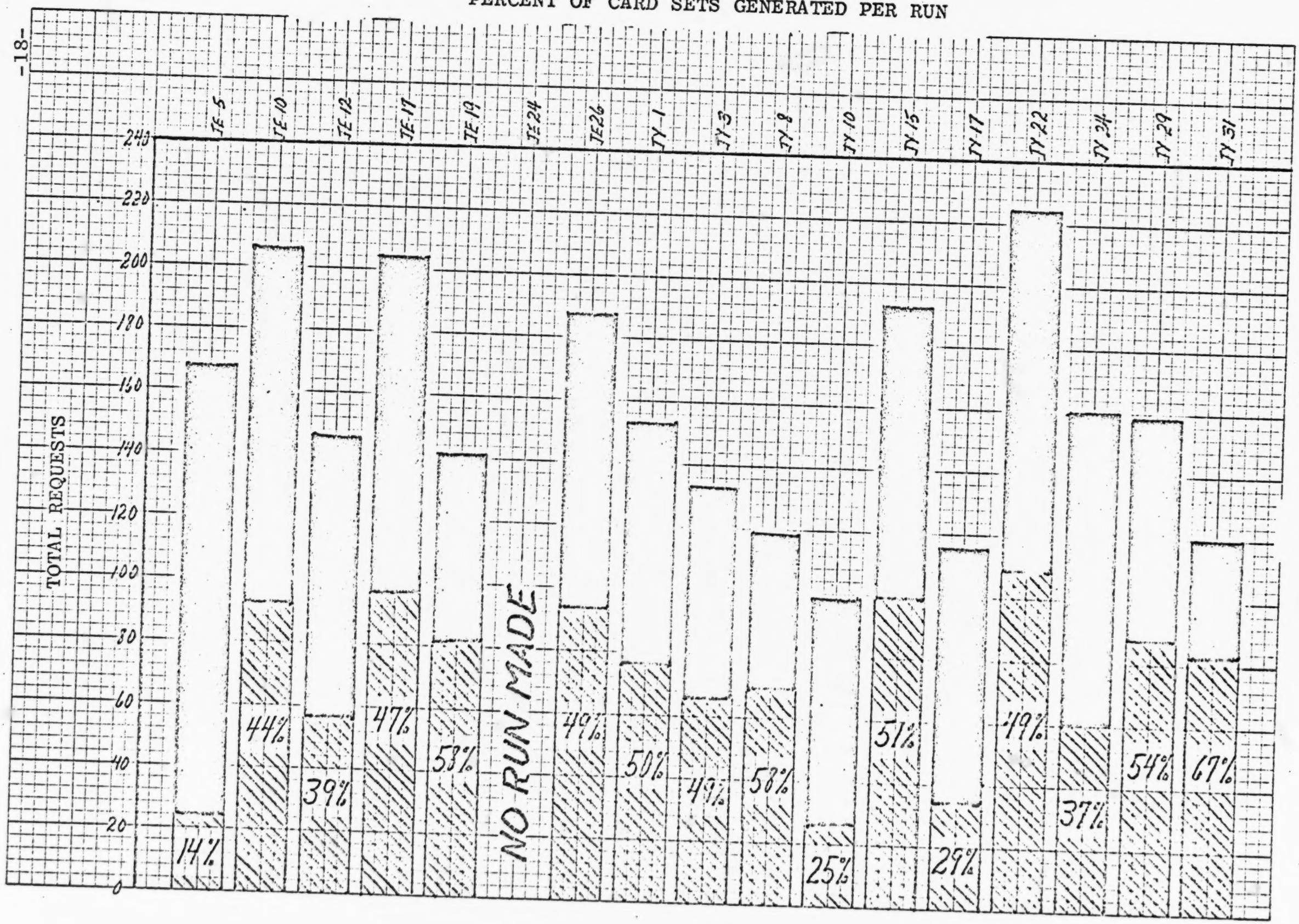
Projected costs of a full scale random access system appear in Appendix A.

2.3 PILOT DEMONSTRATION SUSPENDED JULY 31

Although the majority of the libraries and the members of the Advisory Group thought that the pilot demonstration was well worth continuing, it was finally decided to suspend the operation at the end of July and start it up again when the MARC II system was up and running. A major factor in this decision was that the Library of Congress was discontinuing their MARC I pilot operation. Since no new cataloging data would be added after June 30, the value of continued searching of the tapes to provide cataloging copy for new acquisitions at the NELINET libraries was questionable. Since the project hopes to restart the NELINET pilot operation with the MARC II tapes as soon as possible, the teletypewriters were left in the libraries so that interlibrary loan activities that have developed among the NELINET libraries would not be disturbed.

TABLE IV

PERCENT OF CARD SETS GENERATED PER RUN



▨ = Card sets made

TABLE V

MACHINE RUNNING TIMES

	TIME/REQUEST (SEC.)			TIME/TITLE (SEC.)			TIME/CARD * (SEC.)		
PROGRAM TITLE	JUNE	JULY	COMB. AVG.	JUNE	JULY	COMB. AVG.	JUNE	JULY	COMB. AVG.
Request Validation	3.87	2.85	3.28	9.24	5.94	7.20	0.95	0.66	0.79
Request Sort	2.67	3.04	2.86	6.24	6.36	6.30	0.64	0.71	0.69
Search	22.38	34.86	29.70	53.40	72.60	65.40	5.51	8.22	7.08
Card Production	3.52	4.20	3.91	8.40	8.76	8.64	0.87	0.98	0.93
Punch Labels	1.19	1.71	1.49	2.85	3.56	3.29	0.29	0.40	0.35
Card Formatter	6.24	8.46	7.50	14.94	17.58	16.56	1.54	1.98	1.75
Line Printer	32.70	34.98	33.90	77.40	72.60	74.40	8.04	8.16	8.10
TOTAL	72.57	90.10	82.64	172.47	187.40	181.79	17.84	21.11	19.69

* 9.7 cards per title

TABLE VI

MACHINE RUNNING COSTS *

	COST/TITLE ($)			COST/CARD ($) **		
PROGRAM TITLE	JUNE	JULY	COMB. AVG.	JUNE	JULY	COMB. AVG.
Request Validation	.0767	.0493	.0597	.0078	.0055	.0065
Request Sort	.0518	.0528	.0523	.0053	.0059	.0057
Search	.4432	.6026	.5428	.0457	.0682	.0587
Card Production	.0697	.0727	.0717	.0072	.0081	.0077
Punch Labels	.0236	.0295	.0273	.0024	.0033	.0029
Card Formatter	.1240	.1459	.1374	.0129	.0164	.0145
Line Printer	.6424	.6025	.6175	.0667	.0677	.0672
TOTAL	1.4314	1.5553	1.5087	.1480	.1751	.1632

* All machine running costs are computed at the rate of $30 per hour
** 9.7 cards per title

3. CONVERSION TO THE MARC II BASED SYSTEM

3.1 INTRODUCTION

The MARC I format was barely a reality when the NELINET project began. Now it is passé. Planning how the project should schedule and manage its conversion to the MARC II format consumed much time and effort during this reporting period. Basic to these deliberations were four major considerations:

(1) All future research and development activities should be based on the MARC II standard.

(2) Programming for machines other than the eventual configuration of the NELINET processing center should be kept to a minimum.

(3) The capability to create data files and generate products from them should be developed and demonstrated since such a capability is a necessary and important component of any full scale operation.

(4) The demonstration of catalog support services should be resumed as soon as possible to retain interest and enthusiasm in the project.

At the end of the previous contract, the plan was to convert MARC II tapes into MARC I tapes[1]. These tapes could then be run through the existing card production programs. The reasons for this approach were (1) to delay reprogramming until MARC II was firmly established, (2) to gain more operating experience with the existing programs, and (3) to provide MARC II based card production services as soon as possible.

Deliberations on the method for converting to MARC II began by studying the various approaches that could be taken. Each approach was analyzed in the light of the four major considerations noted above, and evaluated in terms of the total amount of programming required and the long term usefulness of the programs. Converting Library of Congress MARC II format to Library of Congress MARC I format and converting Library of Congress MARC II format to NELINET MARC I format were both rejected because they did not provide for locally inputting and identifying data according to the MARC II standard.

---

[1] Agenbroad, J. E., et al, Systems Design and Pilot Operation of a Regional Center for Technical Processing for the Libraries of the New England State Universities, Final Report CLR-385, Inforonics, Inc., April 5, 1968, Vol. II, p. D-5.

Even though the Library of Congress MARC II format is a "communications format," designed for the exchange of bibliographic data, it was examined to see if it could be used as the internal NELINET format. In the MARC II report[1] it states:

> "It is recognized that each institution may have an individualized local format tailored to its own needs. Many kinds of machines will probably be used. But if an institution is to send or receive data, only a single translation program should be necessary to convert the local format from or to the communications format."

The Library of Congress format was first examined to see how well the data content and its identification suited the needs of the NELINET processing center. This examination was concerned primarily with the treatment of local data. The Library of Congress format specifies separate data fields for three types of local data--system number, call number, and subject headings--and then reserves the 900 block of numbers for local use. The call number field is broken down into three subfields--call number, holding collection code, and number of copies. Since the generation of labels requires additional information--namely the copy numbers (not number of copies) and volume numbers (or designations) of the physical volumes in each location--the Library of Congress MARC II standard could not be used without modification to provide the services NELINET had provided in its MARC I system.

In addition to considering the needs of services already developed, some thought was given to the data file requirements of the future. As a result a technique for distinguishing local data from data assigned at the Library of Congress was developed which provides considerable flexibility in file organization and ease in processing. Using this technique, the distinction that a data field was assigned locally can be applied to any one of the 200 data types that may be present in a MARC II record.

The second aspect of the Library of Congress format examined was the structure of the machine record and how well it suited NELINET's machine processing functions--those already performed as well as those planned for the future. In the Library of Congress MARC II format, the record directory (map) contains the identification tag, the length, and the starting character position of each of the variable fields in the record. The first two character positions of each data field, however, contain indicators

---

(1) Avram, H.D., Knapp, J.F., and Rather, L., The MARC II Format, A Communications Format for Bibliographic Data, Washington, Library of Congress, January 1968, p. 2.

which further identify the data field. This divisioning of field identification into both directory and data field does not take full advantage of the principles of a "mapped" record.

The Library of Congress format, therefore, was not selected as the NELINET internal format. Whether NELINET used the Library of Congress format or developed its own, the need for a Master File Generator would not have been eliminated--one would have been required in either case. The use of the Library of Congress format would have eliminated the need for two other programs: a program to convert Library of Congress MARC II data into NELINET MARC II master file format and a program to convert NELINET MARC II data into the Library of Congress MARC II communications format. This last program is not required for catalog support services but will be used whenever NELINET wishes to send its data to other institutions. Since this is not an immediate need, it is not being considered in this initial MARC II programming effort.

Early plans, therefore, called for writing three major programs: (1) a NELINET MARC II Master File Generator, (2) a program to convert Library of Congress MARC II data into NELINET MARC II format (3) a program to convert NELINET MARC II data into NELINET MARC I format. The output from this last program could then be run through the existing card production programs. In this scheme, two of the three programs required would be of long term value. It also would allow the project designers to think in terms of MARC II codes in all future development work.

Closer study revealed that the conversion of NELINET MARC II into NELINET MARC I would involve a considerable amount of programming. In addition to routines that would convert (or collapse) each MARC II tag and delimiting sequence into its proper MARC I equivalent, a number of special routines would be needed to handle the structural differences in the format. For example, in MARC II, series statements beginning with a pronoun require special processing so that the series added entry will be correct, i.e., so that it will contain the main entry. Also, in MARC II the tracings "Title" and "(Series)" are not in the data as they were in MARC I, but must be generated by the program.

The fragile nature of the existing PDP-1 card production programs was also becoming increasingly obvious. As might be expected, these programs went through many changes during their development and testing. Attempts to further modify these programs for use in another experimental project showed that they would be error prone in operation and awkward to use.

These factors led to second thoughts about the overall advantages of the "quick" approach. When it was learned that the Library of Congress would issue MARC II tapes later than originally

planned, it was decided that this delay could provide the opportunity to develop new card production and card formatting programs that would be MARC II based and of long term value to NELINET.

There were additional reasons to support this decision. Plans were being made to develop these programs on the PDP-10 by using a service bureau. With the PDP-10, a much larger and more powerful machine than the PDP-1, it is possible to design and develop a system which is much more production-oriented than a system using the PDP-1. In the PDP-1 MARC I based system, the size of the programs became so large that they had to be divided into separate programs. Operating such a system with separate programs necessitated a multiple pass operation with tape storage between passes. By using temporary disc storage with the PDP-10 to store the results of each separate functional operation, it is possible to operate a complete set of programs in what appears to be a one pass operation.

The final decision was, therefore, to adopt the long range approach, putting the programming effort into programs that have long term value rather than writing a MARC II to MARC I converter and modifying the other existing programs.

3.2 SYSTEMS DEVELOPMENT AND PROGRAMMING

What started out to be a rapid conversion so that MARC II based services could begin as soon as possible has gradually evolved into a complete new MARC II system. The system is not yet considered to be final and will continue to evolve for some time to come.

The systems planning effort for the MARC II system has certain advantages that were not available to the MARC I systems effort. Among these are:

(1) More complete information. The MARC II system has been described in minute detail by the Library of Congress in their various reports and specifications. Such information was not available when the NELINET MARC I system was designed and this necessitated many changes in the programs. The Library of Congress is to be praised for providing this information in the midst of their own efforts to get MARC II up and running.

(2) The experience with the NELINET MARC I demonstration. During the final weeks of the demonstration the MARC I programs were thoroughly analyzed and the problem reports submitted by the libraries were reviewed.

(3) The experience with another experimental project. The NELINET MARC I system processed current acquisitions in five state university libraries. Trying to use this system to reclassify retrospective holdings in a public library was a rigorous test of it. As a result of this experience a more flexible system can be designed for MARC II.

(4) Familiarity with the existing acquisitions systems in the five libraries. Although no claims are being made that a totally integrated system is being designed at this time, the knowledge of the acquisition systems in the libraries is useful background information since the two systems must be coordinated.

There are five major programs being developed for the MARC II set up. These are described in the following sections. Other auxiliary programs and routines will be written and used along with these programs and other Inforonics programs when the demonstration of services goes into operation.

3.2.1 Master File Generator

The purpose of this program is to create NELINET MARC II master file records from locally keyed input data. In the input record the identification tags are interspersed among the data fields--each data field being preceded by the tag which identifies it and a carriage return. In the NELINET MARC II master file record format, all the data fields are gathered into one part of the record and all the tags are gathered into another part called the directory. Along with each tag in the directory is the starting character position of the data field that the tag identifies. A sort field appears at the front of each record. This sort field contains the Library of Congress card number, if present, and a local systems number.

In addition to structuring the machine record, this program performs a number of functions aimed at facilitating the input tagging, keying, and proofreading operation. Close contact has been maintained with the Library of Congress on this subject, and many

of the techniques used at the Library of Congress are being incorporated into the Inforonics program, including:

(1) The insertion of the delimiter character sequence representing the first subfield in each data field.

(2) The insertion of the delimiter characters, which have been recorded with the tag, into their appropriate place in the data.

(3) The insertion of the fixed field codes for illustration types by having the program scan the illustration subfield in the collation.

(4) The indication that the main entry is the author of the series by scanning the first word of the series statement.

The program will also expand subdivisions of subject headings so that the typist may type just what she sees.

In addition to these techniques the program will have many built-in checks to help catch errors. The tables which specify the processing that is to be performed on each data field indicate whether the data field may occur only once in a record or may be repeated, and whether the data field must be present in a record. The validity of the delimiting character sequences is also checked for each data field.

Although machine detectable errors do not represent the majority of errors made, the complexity of inputting data in the MARC II format warrants using the computer to help the input operation as much as possible. Special checking routines are also being planned for other data fields such as copy number statements and call numbers. This program has been specified and is being programmed.

3.2.2 LC MARC II to NELINET MARC II Converter

The purpose of this program is to convert the MARC II tapes distributed by the Library of Congress into tape records that are in the NELINET MARC II master file format. In the Library of Congress format the tag identifying each field is in the directory; the indicators which further identify each field occupy the first two positions of the data field. The Inforonics program converts, by algorithm, the Library of Congress tag and indicators into an 18 bit configuration which identifies the data field completely. This 18 bit configuration is the tag which appears in the NELINET MARC II master file directory.

In addition this program moves the data in the Library of Congress leader, which cannot be regenerated automatically, into the variable fixed field and converts the character codes into master file character codes. This program was checked out with the Library of Congress test tape in late September.

3.2.3 Search/Merge

Input to the program will be any or all of the following:

(1) New MARC II records received from the Library of Congress.

(2) New NELINET MARC II records keyed at Inforonics.

(3) Requests for catalog cards and labels received from the libraries.

The Search/Merge Program will:

(1) Match this input file against the NELINET Master File.

(2) Output a new Master File which will contain the old file plus the new records from the Library of Congress, the new locally keyed records, and the unfulfilled requests. These unfulfilled requests will be matched against future shipments from the Library of Congress.

This program is in the process of being specified.

(3) Output a file of records ready for card production processing (new locally keyed records as well as requested records that were found in the Master File or in the new records from the Library of Congress).

(4) Output "not-found" messages for all requests that were not found in the Master File.

3.2.4 Card and Label Production

The card and label production program accepts NELINET MARC II master file records and generates for each input record a complete set of card and label output records. Both input and output records are on disc. Input records contain both bibliographic

data (e.g., main entry, title, etc.) and local data (e.g., location symbols, copy numbers, etc.). Input may be data keyed at Inforonics, (the output of the Master File Generator) or data keyed at the Library of Congress in combination with the local data transmitted by the libraries via the teletypes.

Contained in this program is a profile for each library. This profile consists of information in coded form about each library's processing specifications and includes the following information:

(1) A table of valid branch, department, and special shelf locations giving the card requirements (the number of main entries, added entries, and shelf list cards) for each location.

(2) An indicator for Selin label production.

(3) An indicator for book card production.

(4) Oversize determinations.

(5) Oversize symbols.

(6) Conventional title treatment (print all conventional titles, print only those conventional titles that are printed on Library of Congress printed cards, or do not print any conventional titles).

(7) Main entry as subject treatment indicator.

In processing each record the program will examine the library's profile and perform the operations specified. In addition to the above seven characteristics being programmed at present, provision is being made to accommodate additional profile information for increased future capabilities.

The card and label production programs perform a number of processing functions on bibliographic and local data including the following:

(1) Generation of overprint headings from tracings, titles, and series statements.

(2) Generation of tracings for title and series entries when the overprint headings are taken from the title and series statements.

(3) Generation of the appropriate number of main entries, added entries, subject entries, and shelf list cards from the profile and tracings data.

(4) Generation of the appropriate Arabic or Roman numeral to be printed before each tracing.

(5) Breaking up of the Library of Congress call number string into segments which can be printed in the margin of the cards and on the labels.

(6) Generation of a record for each label from the summarized statement of copies and volumes.

(7) Addition of the library's location symbols (including oversize when appropriate) to the classification-book number to make a complete call number.

This is the "work-horse" program of the card production system since it performs the major processing routines. This program has been specified and is being programmed.

3.2.5 Card Formatter

The purpose of this program is to format the data for the desired output on the particular output device that is to be used. Its major functions include:

(1) Horizontal and vertical positioning of each data field.

(2) Determining where line breaks should occur.

(3) Right-justifying data fields when necessary.

(4) Converting master file character codes into the character codes required by the output device.

(5) Generating continuation card headers and "continued on next card" messages when necessary.

(6) Truncating overprint headings when they contain more than three lines of data.

This program will be table driven. For each data field a leading and trailing message will be specified. For some data fields the leading message will specify a line number and a character position, e.g., the main entry begins on line 4 in character position 9. Other fields, e.g., imprint, follow a specified number of spaces after the preceding data field.

Since cataloging data is highly conditional in that most data fields may or may not be present, the specification of leading and trailing messages for each element is not simple. In some cases the printing specifications for the first occurrence of a data field are different from the specifications for other occurrences of the field. The first series statement, for example, follows the collation. The second begins on a new line in character position 11. The new card formatting program is being designed to handle the differences in printing format occasioned by (1) the presence of other data fields in the record and (2) the number of occurrences of a particular data field.

The cards printed in the MARC I demonstration were printed at 10 characters to an inch and 6 lines to an inch, the normal output for line printers. This is considerably larger than the type on Library of Congress printed cards and many more cards were thus required. The libraries objected to the resulting increased bulk that was going into their catalogs. In designing the new print format, an attempt has been made to reduce the number of cards required as much as possible. The previous format of:

Main Entry

Title_____

_____

has been changed to

Main Entry

Title_____

_____

This will save two spaces on each line. Also the word "Title" and "Series" in the tracings have been reduced to "T" and "S." This program is being specified.

4. CONCLUSION

In implementing the NELINET cataloging support services in five libraries it was shown that the ideas, design, and development which had existed previously on paper could be translated into an actual operating system. As might be expected, many problems arose. In solving these problems a number of results were achieved. First, the libraries became active participating members of the network contributing to its development, and experiencing in the beginning stages its burden, and in the final stages its benefits. Secondly, it showed that a centrally produced machine readable bibliographic record could be supplemented and adapted to suit the needs of not one but five libraries. Third, it showed the potential of a centrally produced, national standard, machine readable bibliographic record. MARC I was a pilot operation at the Library of Congress. MARC II will cover all current English language monographs.

This experience has been invaluable in designing the MARC II system. The new programs reflecting this experience and being designed for a larger machine should result in a smoothly operating system for a full scale NELINET operation.

APPENDIX A

PROJECTED COSTS OF A FULL SCALE RANDOM ACCESS SYSTEM

A.1 INTRODUCTION

The operating costs which have been collected for the experimental system provide the information needed as a basis for a projection of future systems' costs. The purpose of this appendix is to analyze each of the processing functions in the experimental system and to estimate what the cost of that function would be in a full scale random access system. The primary assumption made in the projection is that the equipment used in the future system is fully loaded on a one shift basis. Thus, the costs estimated are minimum, and any project to create a production service will have to plan a method for covering additional costs of startup where the equipment is not fully loaded. The costs estimated are for card and label production only.

A.2 COST ANALYSIS

There are three primary cost categories measured in the experiments: computer processing; telecommunications; and labor, materials, and miscellaneous equipment costs. In our calculation of these costs we have computed to the nearest cent. In addition to these categories of costs there are two others which should be taken into account in an estimate of processing costs of a future system. These are computer overhead costs and systems maintenance costs.

A.2.1 Computer Processing

There are several computer programs which were used in the experimental operation. The experimental cost of operation and the estimated cost of operation in future systems are as follows.

A.2.1.1 Request Validation

The request validation program processes teletype requests from the libraries. The program detects data format errors in teletype transmission, and also rearranges the request into a machine format suitable for subsequent processing. The program costs $.0597 per title to operate. In a future system two factors will reduce this cost. The first factor assumes that the experimental title hit rate of 44.8% will increase to 85%. At this hit rate the request processing time per title found will decrease to $.032 per title found. This estimate is made by computing the time per title found and multiplying the computer rental of $30 per hour.

$$(3.28 \text{ sec/request}) \times (1 \text{ request}/.85 \text{ titles found}) \times (\$30/3600 \text{ sec}) = \$.03/\text{title found}$$

The second factor is that when the requests are entered directly to the computer they will be time shared so that the computer is not slowed by the input paper tape reader. We have assumed no actual saving due to this second factor in our calculations because the method used in the new system will be so different that accurate estimating is difficult. However, it is a safe assumption that the cost of request validation will be further decreased due to this second factor.

A.2.1.2 Request Sort

This program places the requests in order by LC card number so they can be efficiently searched by the batch processing program. This program will not be needed in a random access system so its cost will not be incurred.

A.2.1.3 Search

The search function is presently performed by the experimental system at a cost of $.54 per title. In the random access system,* a simple LC card number search is estimated to take .6 seconds. The cost of this search is estimated to be $.060 assuming a fully loaded one shift operation with a machine rental cost of $360/hour. This is a conservative cost estimate because the .6 seconds is memory access time only, the central processor of the computer being free to do other tasks. The estimate considers the whole computer utilized for .6 seconds, not just the memory.

A.2.1.4 Card Production

This program processes the titles found by the search program to produce a magnetic tape used to print catalog cards and Selin labels. Its cost of operation is $.072 per title. This cost will remain unchanged in the random access system because the cost is made up of central processor time primarily. The central processor time used will be less in the larger computer of the new system but its rental cost will be proportionately more; so the final cost of the processing function is estimated to be the same as at present.

A.2.1.5 Punch Labels

This is an output program which prepares Selin labels. It uses a magnetic tape to prepare a paper tape which will operate a tape typewriter fitted with a Selin labeling attachment. The cost of this function in a future system will remain the same, namely $.03 per title.

A.2.1.6 Card Formatting

This program prepares magnetic tape card images from the card production magnetic tape suitable for printing with an upper and lower case line printer. The present cost of this program, $.14 per title, is expected to decrease by 50% using the new system because 50% of the cost is due to input and output operations which will not be incurred in the future system. The future cost, therefore, is calculated to be $.07.

A.2.1.7 Line Printing

The present line printer program prints one card at a time on an IBM 1403 printer at approximately four lines per second, and costs $.62 per average title of ten cards. The future system

---

*Nugent, William R., Development of Computer Programs and Pilot Operation of a Center to Perform Technical Processing for the New England University Libraries, Quarterly Progress Report of CLR-385, Inforonics, Inc., November 20, 1967.

will use a Data Products line printer and will print cards two across at eight lines per second. This printer costs $20/hour on a fully loaded one shift basis. The printer rental cost, therefore, is 1/6 of the present cost or $.10 per title.

A.2.1.8 Total Computer Processing Costs for Card Production

The total estimated cost of operation of the experimental programs to produce catalog cards and labels on a future random access system is below.

Request Validation	.03
Search	.06
Card Production	.07
Punch Labels	.03
Card Formatting	.07
Line Printing	.10
	.36

A.2.2 Telecommunication Costs

The second category of costs are telecommunication costs. The experimental teletype costs for Vermont's use of a full scale operation are calculated below. Vermont was taken as a typical library because their use in the test was consistent and steady. The following is a calculation of message costs.

30 library requests take 5½ minutes to transmit
6 minutes of telephone line message units cost $1.30
$1.30/30 requests = $.04 per request

$.04 is the cost of one way transmission. With a hit rate of 85%, 15% of the requests must be transmitted back to the library.

.15 x .04 = .01

The total message costs, therefore, are $.04 + $.01 = $.05.

The fixed cost of the teletype is $107 per month. In June, Vermont received 137 titles. The experimental fixed costs per title therefore are:

$107/month ÷ 137 titles-found/month = $.78 per title-found

It is estimated that Vermont's usage of a full scale system would be 10 times that of the experiment. The fixed cost per title found, therefore, would be $.078, or $.08 rounded off to the nearest cent.

The total message cost per title in a future system, therefore, is computed to be:

$.08 + $.05 = $.13

A.2.3 Material, Labor, and Miscellaneous Equipment Costs

Material, labor and miscellaneous equipment are a third category of cost. Costs of the items in the future will be the same as the cost of the items in the experiment except for the card stock which will drop from $.13 per title to $.08 per title (ten cards).

These costs per title are as follows:

Selin label tape printout typewriter rental	$.01
Catalog card stock	.08
Selin labels	.03
Mailing cost	.03
Card Cutter rental	.01
Labor for card and label handling and packaging	.10
	$.26 per title

A.2.4 Total Estimated Future Costs of Card Production Processing Functions Performed in Experimental System

The preceding estimated costs are totaled as follows:

Computer costs	$.36
Telecommunication costs	.13
Material, labor and miscellaneous equipment costs	.26
	$.75 per title

A.2.5 Computer Overhead Costs

In addition to the processing performed in the experiment the future system will have to perform the functions of accounting, billing, and the monitoring of error conditions. These functions are estimated to absorb computer time at an "overhead" ratio of 80%. Therefore, additional computer costs should be included in the amount of:

.8 x 36 = $.29

A.2.6 System Maintenance

After the initial system has been developed its continued improvement should be budgeted as a cost. Such a cost will be proportional to the computer time used for the system function being improved. It is estimated that such maintenance should equal 20% of the basic processing computer cost which is:

.2 x ($.36 + $.29) = $.13

A.2.7 Total Computer Cost

The total computer cost considering these additional functions not included in the experiment is:

Card production computer cost	$.36
Computer "overhead" for automatic accounting and system monitoring	.29
Cost of continued program maintenance and systems improvement	.13
	$.78

A.2.8 Total Cost

The total cost in a future system of the card production services provided in the experimental system is, therefore, estimated to be:

Computer cost	$.78
Telecommunications costs	.13
Material, labor and miscellaneous equipment costs	.26
	$1.17